

APPELLANTS' BRIEF ON APPEAL UNDER 37 C.F.R. §41.37  
U.S. Application Serial No. 10/042,192  
Attorney Docket No. 042846-0312966

7340  
AT  
\$



**IN THE UNITED STATES PATENT AND TRADEMARK OFFICE**

IN RE PATENT APPLICATION OF: Thangaraj Veerappan  
SERIAL NO.: 10/042,192  
FILING DATE: January 11, 2002  
ATTORNEY DOCKET NO.: 042846-0312966  
CONFIRMATION NO.: 6876  
ART UNIT: 2178  
EXAMINER: Jonathon D. Schlaifer  
FOR: SYSTEM AND METHOD FOR DETERMINING A DOCUMENT LANGUAGE AND  
REFINING THE CHARACTER SET ENCODING BASED ON THE DOCUMENT  
LANGUAGE

---

**APPEAL BRIEF UNDER 37 C.F.R. §41.37**

**Mail Stop Appeal Brief - Patents**

Commissioner for Patents  
P.O. Box 1450  
Alexandria, VA. 22313-1450

Dear Sir:

Further to the Notice of Appeal filed on August 25, 2005 and the Notice of Panel Decision from Pre-Appeal Brief Review mailed October 5, 2005, Appellants respectfully submit an Appeal Brief pursuant to 37 C.F.R. §41.37.

The Director is authorized to charge the \$500.00 fee for filing an Appeal Brief pursuant to 37 C.F.R. §41.20(b)(2). The Director is further authorized to charge any additional fees that may be due, or credit any overpayment of same to Deposit Account No. 033975 (Ref. No. 042846-0312966).

11/08/2005 SZEWDIE1 00000141 033975 10042192

01 FC:1402 500.00 DA

**REQUIREMENTS OF 37 C.F.R. §41.37**

**I. REAL PARTY IN INTEREST**

The real party in interest is International Business Machines Corporation.

**II. RELATED APPEALS AND INTERFERENCES**

Appellants are aware of no related appeals or interferences.

**III. STATUS OF CLAIMS**

Pending: Claims 1-40 are pending.

Cancelled: No claims are cancelled.

Rejected: Claims 1-40 stand rejected.

Allowed: No claims have been allowed.

On Appeal: The rejection of claims 1-40 under  
35 U.S.C. § 103(a) is appealed.

**IV. SUMMARY OF CLAIMED SUBJECT MATTER**

One aspect of the invention relates to a system and method that determine a language in which an electronic document is created. *See* the specification at page 2. After receiving an electronic document, the system and method identify a character set in which the text of the electronic document has been created. *See id.* at page 3. Based on the identified character set(s), a set of potential languages that correspond to the identified

character set(s) may be determined, wherein the set of potential languages include languages in which the electronic document may have been created in. *See id.* The number of potential languages in the set of potential languages may be decreased based on further processing. *See id.*

In one embodiment, at least one electronic document that includes a character string may be received, wherein characters in the character string are represented in at least one of a plurality of character sets corresponding to an undetermined language. *See id.* at pages 4 and 5. The at least one electronic document may be received by a receiving module (e.g., electronic document receiving module 602 of FIG. 9). *See id.* at page 17. The recitation of “receiving means” in the claims may refer to *at least* the electronic document receiving module 602 of FIG. 9

In one embodiment, at least a portion of the character string may be evaluated by comparing each of the characters in the portion of the character string to a plurality of predetermined candidate character sets to determine one or more matches between the plurality of predetermined candidate character sets and the characters in the portion of the character string. *See id.* at pages 9 and 10. The evaluation of the at least a portion of the character string may be performed by a character set identification module (e.g., character set encoding identification module 604 of FIG. 9). *See id.* at page 17. The recitation of “evaluating means” in the claims may refer to *at least* the character set encoding identification module 604 of FIG. 9.

In one embodiment, whether one or more character sets that match the characters in the portion of the character string correspond to one or more supported languages may be determined. *See id.* at page 16. This determination may be made by a determining module

(e.g., language determining module 608 of FIG. 9). *See id.* at page 17. The recitation in the claims of “determining means” may refer to *at least* the language determining module 608 of FIG. 9.

In one embodiment, one or more supported languages in which the electronic document is created may be identified based on a determination that the one or more character sets that match the characters in the portion of the character string correspond to one or more supported languages. *See id.* at page 16. This identification may be performed by an identifying module (e.g., character group identification detecting module 614 of FIG. 9). *See id.* at pages 17 and 18. The recitation of “identifying means” may refer to *at least* the character group identification detecting module 614 of FIG. 9.

**V. GROUND OF REJECTION TO BE REVIEWED ON APPEAL (35 U.S.C. § 103).**

Claims 1-4, 6-14, 16-24, 26-34, and 36-40 stand rejected under 35 U.S.C. 103(a) as allegedly being unpatentable over U.S. Patent No. 6,252,671 to Peng *et al.* (“Peng”) in view of U.S. Patent No. 6,104,381 to Watanabe *et al.* (“Watanabe”).

Claims 5, 15, 25, and 35 stand rejected under 35 U.S.C. 103(a) as allegedly being unpatentable over Peng in view of Watanabe, and in further view of U.S. Patent No. 6,167,369 to Schulze (“Schulze”).

**VI. ARGUMENT**

**A. CLAIMS 1-4, 6-14, 16-24, 26-34 AND 36-40**

The rejection of claims 1-4, 6-14, 16-24, 26-34 and 36-40 as being unpatentable over Peng in view of Watanabe must be overruled *at least* (1) because Peng and Watanabe,

both alone and in combination, do not teach or suggest all of the features of the claimed invention, and (2) because there is no proper motivation for combining Peng and Watanabe.

1. *Peng and Watanabe do not teach or suggest all of the features of the claimed invention.*

Claim 1 recites *inter alia*, “evaluating at least a portion of the character string by comparing each of the characters in the portion of the character string to a plurality of predetermined candidate character sets to determine one or more matches between the plurality of pre-determined candidate character sets and the character string.” Peng and Watanabe do not teach or suggest this feature. Independent claims 11, 21, and 31 include similar subject matter, among other things. Claims 2-4, 6-10, 12-14, 16-20, 22-24, 26-30, 32-34 and 36-40 depend from corresponding ones of claims 1, 11, 21, and 31, and therefore include subject matter similar to the recitation provided above, as well as additional features.

Peng appears to teach an apparatus that downloads a font available on a computer to a printer that supports a page description language such as PostScript (see Peng at col. 2, lines 15 – 19). In Peng, information related to a font, such as language information or character information, is already known, and is included in font data transmitted between the computer and the printer in the page description language. Thus, Peng fails at least to teach or suggest deriving such information from the actual characters in the font.

The Examiner contends that Peng “identifies character set encoding in the font documents in col. 5, lines 40-65,” and that this character set encoding is *analogous to* evaluating at least a portion of the character string by comparing each of the characters in the portion of the character string to a plurality of predetermined candidate character sets

to determine one or more matches between the plurality of pre-determined candidate character sets and the character string. *See* the Final Office Action at paragraph number 6. This is incorrect and unsupported.

Instead, the cited passage appears to describe a method step (step 138) of the method illustrated in FIG. 3 at which encoding character set data is sent to a printer. *See id.* at col. 4, lines 57-59; and col. 5, lines 44 and 45. The encoding character set data may be of any one of three separate encoding types used to transfer the font data to the printer, which include: (1) a predefined encoding type, (2) a standard encoding type, and (3) a byte encoding type. *See id.* at col. 4, lines 59-63; and col. 5, lines 51-65. The type of encoding character set data may depend on types of characters included in the font being downloaded (*e.g.*, English-language characters, non-English-language characters, *etc.*). *See id.* at col. 5, lines 51-65. However, the cited passage of Peng fails to teach or suggest having no prior knowledge of character set information and performing to a *comparison* between a character string and a plurality of *candidate character sets* to determine one or more matches between the character string and the candidate character sets.

The Examiner acknowledges that Peng is deficient at least for failing “to disclose that the document includes a character string, wherein the characters in the character string are represented in at least one of a plurality of character sets correspond to an undetermined language.” *See* the Final Office Action at paragraph 6. The Examiner relies on Watanabe to overcome this admitted deficiency. Assuming *arguendo*, that there was proper motivation for combining Watanabe with Peng, the proposed combination would still not teach or suggest all of the features of the claimed invention because the Examiner’s characterization of the teachings of Watanabe is incorrect. The Examiner

alleges that Watanabe discloses that documents of unknown languages are processed in strings. However, Watanabe describes processing character strings from a known language, that include either kana or kanji characters. Since kana and kanji characters are both *Japanese* language characters, Watanabe only suggests processing characters strings that include characters corresponding to a *single* known language.

Additionally, Watanabe does not resolve the deficiency of Peng addressed above because, as is the case with Peng, Watanabe does not teach or suggest evaluating at least a portion of the character string by comparing each of the characters in the portion of the character string to a plurality of predetermined candidate character sets to determine one or more matches between the plurality of pre-determined candidate character sets and the character string. For *at least* these reasons, the rejection of claims 1, 11, 21, and 31, as being unpatentable over Peng in view of Watanabe should be overruled.

2. *There is no proper motivation to combine Peng and Watanabe.*

The combination of Peng and Watanabe is improper because the motivation for combining the references provided by the Examiner consists of impermissible “hindsight” reasoning, and because no suggestion to combine the references exists in the prior art.

In order for the proposed combination of Peng and Watanabe to be proper, “there must be some motivation, suggestion, or teaching of the desirability of making the specific combination that was made by the applicant.” *In re Suang-Su Lee*, 277 F.3d 1338, 61 U.S.P.Q.2d 1430 (Fed.Cir. 2002). The Examiner alleges that it would have been obvious to combine Peng with Watanabe “in order to streamline data processing large amounts of data.” This motivation is not found within the references themselves, and it appears that this motivation was concocted by the Examiner using impermissible

“hindsight” to selectively cull components from the prior art to fit the parameters of the claimed invention. *ATD Corp. v. Lydall, Inc.*, 159 F.3d 534, 546, 48 USPQ2d 1321, 1329 (Fed.Cir.1998). Further, it is not apparent from the teachings of the references themselves that combining Peng and Watanabe in the proposed manner would even provide the benefit presented by the Examiner. For *at least* this reason, the rejection of claims 1, 11, 21, and 31 based on the combination of Peng and Watanabe is improper and should be overruled.

B. CLAIMS 5, 15, 25 AND 35

The rejection of claims 5, 15, 25, and 35 as being unpatentable over Peng in view of Watanabe, and in further view of Schulze is improper *at least* because there is no proper motivation for combining Peng and Watanabe with Shulze in the proposed manner.

The Examiner admits that the combination of Peng in view of Watanabe is deficient at least because the proposed combination of Peng and Watanabe fails to teach or suggest that the at least one group of characters is an n-gram, and relies on Schulze for a teaching that n-grams may be used to facilitate probabilistic analysis of whether a language is predominant. *See* the Final Office Action at paragraph 43.

Schulze is apparently drawn to automatically identifying a predominant language of a document based on both probability data and word data. *See* Schulze at the Abstract. The alleged motivation for modifying the combination of Peng and Watanabe is that it would have been obvious to “facilitate probabilistic analysis of whether a language is predominant.” *See* the Final Office Action at paragraph 43. However, even if this were true, the facilitation of probabilistic analysis of whether a language is predominant would not have constituted a motivation, suggestion, or teaching of the *desirability* of the



proposed combination. *See Suang-Su Lee* (emphasis added). Thus, the alleged motivation provided by the Examiner for combining Peng and Watanabe with Shulze is improper.

Further, a proper motivation for making the proposed combination does not exist in the prior art. For *at least* this reason, the rejection of claims 5, 15, 25, and 35 is improper and must be overruled. Claims 5, 15, 25, and 35 are also allowable by virtue of their dependency from claims 1, 11, 21, and 31, respectively, for the reasons provided above.

**VII. APPENDIX**

- A. EVIDENCE (NONE)
- B. RELATED PROCEEDINGS (NONE)
- C. PENDING CLAIMS

The pending claims (claims 1-40) may be found in the Appendix included herewith.

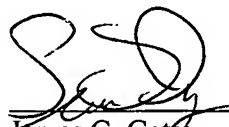
CONCLUSION

For at least the foregoing reasons, Appellant respectfully requests that the rejection of each of pending claims 1-40 under 35 U.S.C. §103(a) be reversed.

Date: November 7, 2005

Respectfully submitted,

By:

 #40203  
James G. Gatto  
for Registration No. 32694

Customer No. 00909

Pillsbury Winthrop Shaw Pittman LLP  
P.O. Box 10500  
McLean, Virginia 22102  
Main: 703-770-7900  
Fax: 703-905-2500

**APPENDIX C.**

**1. (Previously Presented)** A method for determining a language in which a document is created comprising the steps of:

receiving at least one electronic document that includes a character string, wherein characters in the character string are represented in at least one of a plurality of character sets corresponding to an undetermined language;

evaluating at least a portion of the character string by comparing each of the characters in the portion of the character string to a plurality of predetermined candidate character sets to determine one or more matches between the plurality of predetermined candidate character sets and the characters in the portion of the character string;

determining whether one or more character sets that match the characters in the portion of the character string correspond to one or more supported languages; and

identifying one or more supported languages in which the electronic document is created based on a determination that the one or more character sets that match the characters in the portion of the character string correspond to one or more supported languages.

**2. (Previously Presented)** The method of claim 1, wherein the step of determining includes determining that the one or more character sets that match the characters in the portion of the character string correspond to at least two supported languages associated with the electronic document.

**3. (Previously Presented)** The method of claim 2, further comprising the step of comparing at least one group of characters in the portion of the character string to predetermined groups of characters.

4. **(Previously Presented)** The method of claim 3, further comprising the step of detecting at least one identification for the at least one group of characters.

5. **(Original)** The method of claim 3, wherein the at least one group of characters is an n-gram.

6. **(Original)** The method of claim 4, wherein the at least one identification is a bit-flag.

7. **(Previously Presented)** The method of claim 4, further comprising the step of logically ANDing the at least one identification.

8. **(Previously Presented)** The method of claim 7, wherein the step of logically ANDing the at least one identification is repeated until a single identification is determined.

9. **(Previously Presented)** The method of claim 8, further comprising the step of indicating the supported language associated with the electronic document.

10. **(Previously Presented)** The method of claim 9, further comprising the step of identifying a character set associated with the supported language indicated.

11. **(Previously Presented)** A system for determining a language in which a document is created comprising:

receiving means for receiving at least one electronic document that includes a character string, wherein characters in the character string can be represented in any of a plurality of character sets corresponding to an undetermined language;

evaluating means for evaluating at least a portion of the character string by comparing each of the characters in the portion of the character string to a plurality of predetermined candidate character sets to determine one or more matches between the plurality of predetermined candidate character sets and the characters in the portion of the character string;

determining means for determining whether one or more character sets that match the characters in the character string correspond to one or more supported languages; and

identifying means for identifying one or more supported languages in which the electronic document is created based on a determination that the one or more character sets that match the characters in the portion of the character string correspond to one or more supported languages.

**12. (Previously Presented)** The system of claim 11, wherein the determining means determines that the one or more character sets that match the characters in the portion of the character string identify at least two supported languages associated with the electronic document.

**13. (Previously Presented)** The system of claim 12, further comprising comparing means for comparing at least one group of characters in the portion of the character string to predetermined groups of characters.

**14. (Original)** The system of claim 13, further comprising detecting means for detecting at least one identification for the at least one group of characters.

**15. (Original)** The system of claim 13, wherein the at least one group of characters is an n-gram.

16. **(Original)** The system of claim 14, wherein the at least one identification is a bit-flag.

17. **(Original)** The system of claim 14, further comprising logical ANDing means for logically ANDing the at least one identification.

18. **(Original)** The system of claim 17, wherein the logically ANDing means logically ANDs the at least one identification until a single identification is determined.

19. **(Previously Presented)** The system of claim 18, further comprising language indicating means for indicating the supported language associated with the electronic document.

20. **(Previously Presented)** The system of claim 19, further comprising character set identifying means for identifying a character set associated with the supported language indicated.

21. **(Previously Presented)** A system for determining a language in which a document is created comprising:

a receiving module that receives at least one electronic document that includes a character string, wherein characters in the character string can be represented in any of a plurality of character sets corresponding to an undetermined language;

a character set identification module that evaluates at least a portion of the character string by comparing each of the characters in the portion of the character string to a plurality of predetermined candidate character sets to determine one or more matches between the plurality of predetermined candidate character sets and the characters in the portion of the character string;

a determining module that determines whether one or more character sets that match the characters in the portion of the character string correspond to one or more supported languages; and

an identifying module that identifies one or more supported languages in which the electronic document is created based on a determination that the one or more character sets that match the characters in the character string correspond to one or more supported languages.

22. **(Previously Presented)** The system of claim 21, wherein the determining module determines that the one or more character sets that match the characters in the portion of the character string correspond to at least two supported languages associated with the electronic document.

23. **(Previously Presented)** The system of claim 22, further comprising a comparing module that compares at least one group of characters in the portion of the character string to predetermined groups of characters.

24. **(Original)** The system of claim 23, further comprising a detecting module that detects at least one identification for the at least one group of characters.

25. **(Original)** The system of claim 23, wherein the at least one group of characters is an n-gram.

26. **(Original)** The system of claim 24, wherein the at least one identification is a bit-flag.

27. **(Original)** The system of claim 24, further comprising a logical ANDing module that logically ANDs the at least one identification.

28. **(Original)** The system of claim 27, wherein the logically ANDing module logically ANDs the at least one identification until a single identification is determined.

29. **(Previously Presented)** The system of claim 28, further comprising a language indicating module that indicates the supported language associated with the electronic document.

30. **(Previously Presented)** The system of claim 29, further comprising a character set identifying module that identifies a character set associated with the supported language indicated.

31. **(Previously Presented)** A processor readable medium comprising processor readable code that causes a processor to determine a language in which a document is created, the processor readable medium comprising:

receiving code that causes a processor to receive at least one electronic document that includes a character string, wherein characters in the character string can be represented in any of a plurality of character sets corresponding to an undetermined language;

evaluating code that causes a processor to evaluate at least a portion of the character string by comparing each of the characters in the portion of the character string to a plurality of predetermined candidate character sets to determine one or more matches between the plurality of predetermined candidate character sets and the characters in the portion of the character string;

determining code that causes a processor to determine whether one or more character sets that match the characters in the portion of the character string correspond to one or more supported languages; and

identifying code that causes a processor to identify one or more supported languages in which the electronic document is created based on a determination that the one



or more character sets that match the characters in the portion of the character string correspond to one or more supported languages.

32. **(Previously Presented)** The medium of claim 31, wherein the determining code determines that the one or more character sets that match the characters in the portion of the character string identify at least two supported languages in the electronic document.

33. **(Previously Presented)** The medium of claim 32, further comprising comparing code that causes a processor to compare at least one group of characters in the portion of the character string to predetermined groups of characters.

34. **(Original)** The medium of claim 33, further comprising detecting code that causes a processor to detect at least one identification for the at least one group of characters.

35. **(Original)** The medium of claim 33, wherein the at least one group of characters is an n-gram.

36. **(Original)** The medium of claim 34, wherein the at least one identification is a bit- flag.

37. **(Original)** The medium of claim 34, further comprising logical ANDing code that causes a processor to logically AND the at least one identification.

38. **(Original)** The medium of claim 37, wherein the logically ANDing code logically ANDs the at least one identification until a single identification is determined.

39. **(Previously Presented)** The medium of claim 38, further comprising language indicating code that causes a processor to indicate the supported language associated with the electronic document.

40. **(Previously Presented)** The medium of claim 39, further comprising character set identifying code that causes a processor to identify a character set associated with the supported language indicated.